

Integrated Gradients

A	x	Ax	b

Saliency

A	x	Ax	b

Integrated Gradients

A	x	Ax	b

Saliency

A	x	Ax	b

Figure 1: Comparison of Integrated Gradients and Saliency methods for feature attribution. The figure displays 10 rows of results, each corresponding to a different input image (represented by a 5x5 heatmap 'A'). For each row, the input image 'A' is multiplied by a vector 'x' to produce a vector 'Ax', which is then added to a bias vector 'b' to produce the final output. The Integrated Gradients method (left) and the Saliency method (right) are used to compute the attribution for each feature in the input image 'A'. The attribution is visualized as a 5x5 heatmap, where the color scale ranges from -1 (blue) to 1 (red). The Integrated Gradients method shows more localized and precise attribution compared to the Saliency method, which often shows more diffuse and noisy attribution patterns.